



CHARTE POUR UNE DEMARCHE QUALITE

REALISATION DE BASES DE DONNEES ISSUES D'ENQUETES DE TERRAIN



Correspondantes Démarche Qualité:
Aubert Magali – aubert@supagro.inra.fr
Morin Pascale – pascale.morin@cirad.fr

SOMMAIRE

1. Introduction
2. La démarche qualité : cartographie du processus « bases de données issues d'enquêtes de terrain »
3. Spécificités des enquêtes réalisées par internet
4. Elaboration/conception du questionnaire
5. Préparation de la collecte
 - 5.1. Sélection des enquêteurs et des superviseurs
 - 5.1.1. Les enquêteurs
 - 5.1.2. Les superviseurs
 - 5.2. Test des questionnaires
 - 5.2.1. Objectif de la phase test
 - 5.2.2. Prise de contact avec les fournisseurs de données officielles
6. Collecte des données
 - 6.1. Prise de contact pour faciliter l'accès au terrain
 - 6.2. Planification des rendez-vous
 - 6.3. Qualité de la collecte et mode de rémunération
7. Saisie et validation des données
 - 7.1. Création des bases de données
 - 7.2. Validation des données

8. Gestion des dossiers relatifs aux étapes du processus du cycle de vie des données

8.1. Base de données

8.2. Programme

9. Sauvegarde et archivage

10. Accès et diffusion des données

10.1. Anonymisation

10.2. Confidentialité des données au sein du groupe de travail

10.3. Déclaration CNIL

10.4. Mise à disposition

1. Introduction

La démarche qualité assure que tous les moyens ont été mis en œuvre pour que les données mobilisées par les chercheurs soient fiables, traçables tout au long de leur cycle de vie, reproductibles, pérennes et transférables. Elle s'inscrit dans une logique de recommandations et de préconisations. Elle permet notamment de répondre aux contraintes croissantes exigées lors des publications.

La démarche qualité appliquée aux données en sciences économiques a été initiée par l'INRA et plus précisément par le département SAE2. Dans ce cadre, une charte, un guide de bonnes pratiques ainsi qu'un pense-bête/auto-évaluation ont été réalisés (http://www.agrotic.org/blog/wp-content/uploads/2013/04/Guide_Demarche_Qualite_Gestion_Prep_Donnees.pdf).

Cette réflexion s'inscrit dans le cadre de données mises à disposition par les services fournisseurs tels que le SSP (Service de la Statistique et de la Prospective).

Les bases de données dites ad hoc, c'est-à-dire construites par les chercheurs, n'entrent pas dans le processus tel que défini précédemment. L'UMR MOISA a ainsi constitué un groupe de réflexion pour pallier ce manque. Afin de prendre en compte la richesse de l'unité, ce groupe est volontairement constitué de chercheurs, d'ingénieurs, de doctorants et d'administratifs¹.

L'UMR MOISA participe à la réalisation de nombreuses bases de données ad hoc, tant au Nord qu'au Sud. De ce fait, il existe une grande diversité en termes de méthodes et d'approches mobilisées. Les partenariats à développer sont par ailleurs à appréhender différemment selon le contexte dans lequel la base de données est construite. Cette hétérogénéité des contextes est à coupler avec la diversité des financements (public versus privé) et donc des contraintes d'exploitation des données (confidentialité, droits de diffusion...). Enfin, les bases issues d'enquêtes de terrain peuvent relever d'analyses qualitatives ou quantitatives; elles sont réalisées par des chercheurs, des doctorants ou des stagiaires. De par ces spécificités, l'UMR MOISA peut contribuer, de façon significative, à la réflexion sur la démarche qualité pour les bases de données dites ad hoc. Elle collabore ainsi à une réflexion naissante menée au CIRAD, sur la démarche qualité de telles bases de données.

¹ Le groupe est constitué de M. Aubert, P. Morin ainsi que de C. Bignebat, P.M. Bosc, C. Boué, E. Bouquet et S. Dury.

Pour relever d'une démarche qualité, la création de données ad hoc nécessite de mettre en œuvre une réflexion sur :

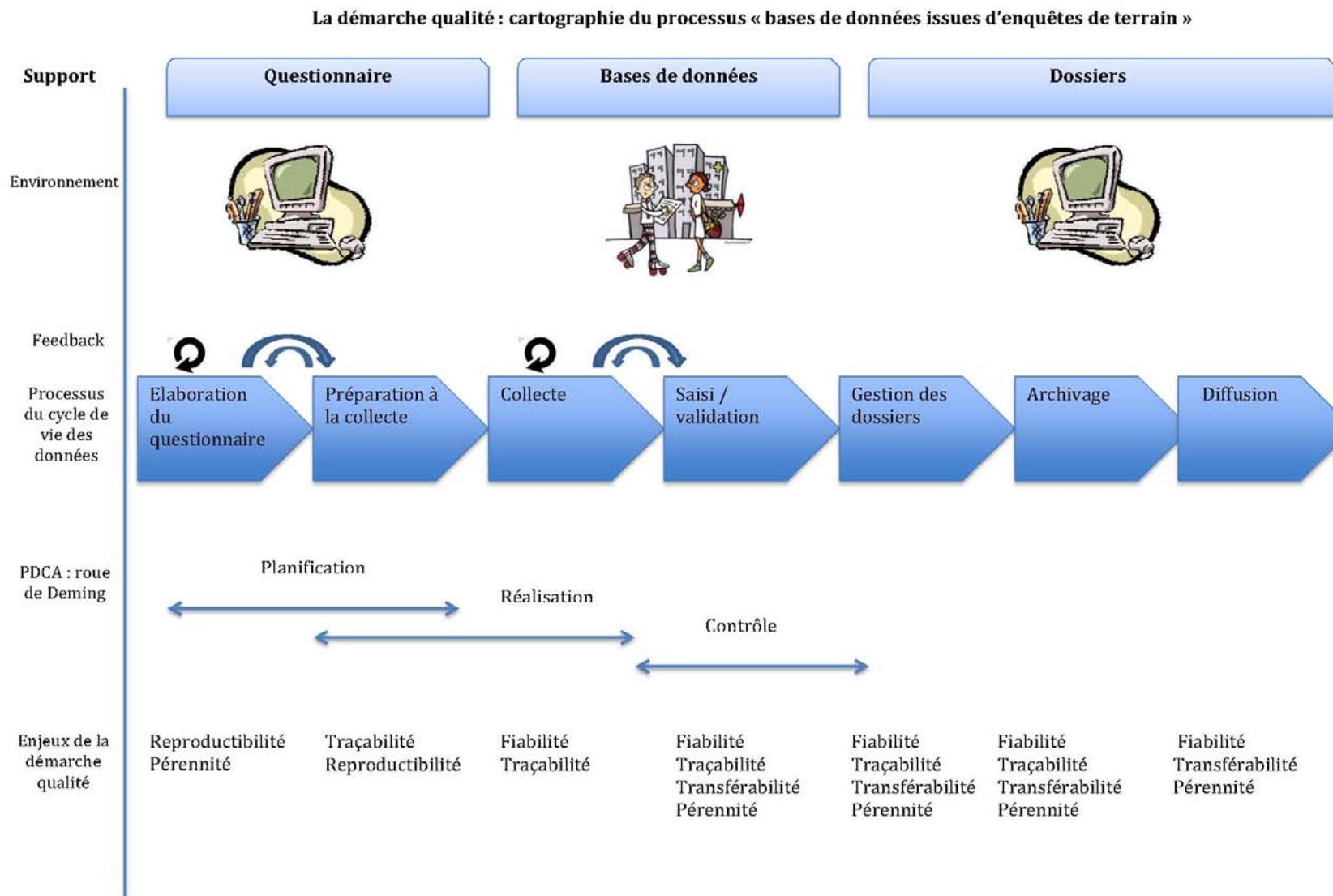
- l'élaboration / la conception du questionnaire
- la préparation de la collecte
- la collecte des données
- la saisie
- la gestion des dossiers relatifs aux étapes du processus du cycle de vie des données
- l'archivage
- la mise à disposition

La démarche qualité est un processus dynamique et interactif dans la mesure où chaque étape de ce processus dépend des autres étapes. Chacune permet de construire la suivante et d'améliorer et enrichir la précédente. En ce sens, elle répond à la logique de la roue de Deming, représentée ci-dessous, et qui souligne l'importance de l'interaction entre la planification, la réalisation, le contrôle et l'amélioration du processus.



La présente charte précise chacune des étapes mentionnées précédemment et spécifie les recommandations aux chercheurs pour la réalisation de bases de données ad hoc. En fin de charte, un pense « pas » bête permet à chacun de faire un bilan des points incontournables à prendre en compte lors de la réalisation d'enquêtes de terrain. Cette grille permet de se positionner et d'établir un degré de réalisation des critères nécessaires à la mise en œuvre d'une démarche de qualité.

2. La démarche qualité : cartographie du processus « bases de données issues d'enquêtes de terrain »



Source : Cartographie réalisée par les auteurs

3. Spécificités des enquêtes réalisées par internet

Les bases de données considérées dans ce document sont des bases issues d'enquêtes de terrain. Les enquêtes réalisées par internet ne sont pas prises explicitement en compte. Toutefois, la plupart des recommandations faites peuvent s'y appliquer. L'UMR MOISA réalise peu d'enquêtes *via* internet. Avec le développement de plateformes collaboratives telles que silverpeas, ce mode d'enquête se développe.

Plusieurs avantages sont liés à la mise en œuvre de ce mode d'enquête : rapidité, gain économique, périmètre élargi des enquêtés.

En contrepartie, plusieurs limites peuvent être observées :

- dans la mesure où il n'y a pas d'enquêteur, il est recommandé :
 - de préciser en introduction : la thématique et la finalité du questionnaire
 - d'introduire des questions qui confirment que l'enquêté correspond à la population ciblée
 - que les enquêtes soient quantitatives
 - que les questions soient simples et fermées
 - que le questionnaire soit court afin que l'enquêté aille jusqu'à son terme.
- l'échantillon
 - les répondants sont, de fait, des internautes. Les personnes interrogées ne sont pas représentatives de la population française (les personnes âgées, les personnes n'ayant pas accès ou l'habitude d'internet ...)
 - les répondants peuvent être sollicités dans le cadre de « panels » c'est-à-dire qu'ils font partie d'une base de données à laquelle ils ont accepté de souscrire en contrepartie d'une rémunération. Le biais de sélection est à prendre en compte
 - du fait du biais d'échantillonnage, il est recommandé de redresser ex-post les résultats obtenus

Dans la mesure où des données individuelles sont recueillies, une déclaration auprès de la CNIL est, selon les cas, obligatoire (Cf. 10.3).

4. Elaboration / conception du questionnaire

L'élaboration / conception du questionnaire est une étape clef qui conditionne la qualité de la recherche menée. La question de recherche doit y être omniprésente et l'ensemble des informations recherchées doit s'y rattacher. Chacune de ces informations doit être justifiée du point de vue théorique et / ou empirique.

De la même façon, il est primordial d'anticiper l'interprétation des résultats attendus. De ce fait, une réflexion sur la représentativité et donc sur les méthodes d'échantillonnage à mettre en œuvre est à discuter dès l'élaboration du questionnaire.

La nature des traitements souhaités doit également être anticipée. La contrainte relative au nombre d'observations diffère en effet selon que les traitements soient qualitatifs ou quantitatifs. Dans ce dernier cas, il est primordial d'anticiper les possibles non réponses qui affectent la qualité des traitements réalisés et donc des résultats obtenus.

La question de recherche nécessite d'obtenir des informations centrales pour lesquelles la validité des réponses est essentielle. Afin d'avoir des informations fiables, il convient de poser plusieurs questions induisant la même réponse pour la révéler et en valider la pertinence. La structure du questionnaire doit être telle que ces informations croisées n'apparaissent pas comme redondantes pour l'enquête.

De la même façon, l'unité d'analyse doit être clairement identifiée tout au long du questionnaire. Si plusieurs unités sont mobilisées, chacune doit apparaître dans une partie ou sous-partie du questionnaire afin de lever toute ambiguïté pouvant se traduire par des réponses non appropriées ou des réponses erronées. Ainsi, un questionnaire portant sur une étude des modes de consommations devra clairement dissocier une partie relative aux consommateurs de celle relative aux produits consommés.

Il est indispensable de faire un bilan sur les 3 contraintes majeures, avant toute réalisation de bases de données ad hoc.

- La première contrainte est une contrainte financière. Au-delà des coûts directs liés à la logistique induite par les enquêtes de terrain (paiement des enquêteurs, frais de déplacement, frais de traduction...), certains coûts se révèlent non négligeables et

doivent alors être correctement anticipés. Il s'agit notamment du temps de formation des enquêteurs (Cf. Collecte des informations).

- La deuxième contrainte est liée à la durée du projet.
- La troisième contrainte est relative à la longueur et à la durée du questionnaire. Un arbitrage s'impose entre longueur du questionnaire et nombre d'enquêtes. Il s'agit non pas de discuter la pertinence du questionnaire mais sa durée au regard du nombre de questionnaires souhaités. En effet, il convient de respecter certaines règles telles que la durée du questionnaire et de limiter le nombre d'enquêtes quotidiennes. Les chercheurs de l'unité estiment qu'au-delà d'1h30, les réponses apportées par les enquêtés sont moins précises. De la même façon, ils recommandent de limiter le nombre d'enquêtes quotidiennes à 4 par jour.

Un sondage auprès des doctorants de l'unité en cours de réalisation permettra de mesurer l'importance de cette contrainte. Dans le cadre de ce sondage, il est demandé plus précisément aux doctorants ayant réalisé des enquêtes de terrain dans le cadre de leur doctorat, de renseigner le tableau suivant :

Nb de questionnaire	Nb de questionnaire / jour	Heure / questionnaire	Durée de l'enquête (en mois)	Temps de préparation du questionnaire	Temps de saisie, contrôle, correction	Nb d'enquêteurs (dont le doctorant)	Nb de superviseur (dont doctorant)	Nb de passage

Afin d'obtenir des données fiables, il est indispensable d'utiliser un vocabulaire adapté et compréhensible. Les termes spécifiques à nos domaines nécessitent la rédaction d'un glossaire réalisé, tout comme le questionnaire, en collaboration avec l'ensemble des partenaires impliqués.

Dans le cas où les enquêtes sont réalisées dans des pays non francophones, une double traduction est recommandée. Il s'agit de traduire une première fois du français vers la langue de destination avant de traduire une seconde fois de la langue de destination vers le français. Riche d'une telle expérience, les chercheurs de l'unité conseillent fortement d'appliquer le principe de la « double traduction » sans perdre de vue que celle-ci à un coût non négligeable à anticiper. De la même façon, il est conseillé d'utiliser les noms latins de produits, par exemple, afin de lever toute ambiguïté lors de la traduction.

Une fois ces contraintes et précautions prises en compte, il est préconisé de présenter le questionnaire à d'autres chercheurs de l'unité qui ont travaillé sur des thématiques proches

ou sur le terrain considéré. Une fois validé en interne, il est indispensable de réaliser des enquêtes tests auprès de la population cible afin de le valider empiriquement. Ces étapes sont également sources de coûts qu'il est important de prendre en compte.

Comme mentionné précédemment, une des spécificités de l'UMR MOISA est d'avoir une expertise de terrain, qui se traduit notamment par la réalisation de nombreuses enquêtes. De ce fait, et dans le cadre de la démarche qualité initiée, un objectif à terme est de créer une bibliothèque des questionnaires réalisés en les différenciant par thématique et pays enquêté. Cette bibliothèque permettra de répondre à l'objectif de reproductibilité et pérennité des données. Une attention particulière devra toutefois être menée sur la gestion de cette bibliothèque. Elle devra avoir comme objectif d'enrichir une réflexion en cours et non d'avoir pour fin une simple mise à disposition des questionnaires existants. L'objectif étant que les expériences des uns servent au plus grand nombre.



Ce qu'il faut retenir

Elaboration / conception d'un questionnaire

- Veiller à l'omniprésence de la question de recherche
- Adapter la taille de l'échantillon au traitement à réaliser
- Croiser les informations pour révéler / confirmer les informations pertinentes
- Identifier l'unité d'analyse
- Trois contraintes majeures
 - financière
 - durée du projet
 - longueur et durée du questionnaire
- Utiliser un vocabulaire adéquat
 - rédaction d'un glossaire
 - double traduction
- Valider les questionnaires
 - en interne
 - sur le terrain

5. Préparation de la collecte

La préparation à la collecte s'effectue sur le terrain et doit permettre de sélectionner les enquêteurs et les superviseurs, de tester les questionnaires et de prendre contact éventuellement avec les fournisseurs de données officielles.

5.1. Sélection des enquêteurs et des superviseurs

Lors de la préparation de la collecte, qui doit idéalement coïncider avec la période test des premiers questionnaires, les superviseurs et enquêteurs sont recrutés de manière rigoureuse.

5.1.1. Les enquêteurs

En fonction de la question traitée, il peut être plus approprié de recruter au niveau local, ou pas. Recruter des enquêteurs locaux peut permettre de pallier certains problèmes (langue, connaissance du contexte, ...). Toutefois, pour garantir la confidentialité des réponses, il peut être préférable de recruter des enquêteurs non locaux. Cette réflexion doit être menée en amont des enquêtes test pour que celles-ci soient correctement appréciées. Dans tous les cas, les enquêteurs doivent connaître le terrain et savoir s'adapter. L'enquêteur devra avoir des qualités relationnelles dans la mesure où certains enquêtés ne savent ni lire ni écrire.

Lors de cette étape d'identification / sélection des enquêteurs, il est souhaitable de prendre contact avec les chercheurs qui ont déjà réalisé des enquêtes dans le contexte considéré. Ces derniers peuvent faire bénéficier de leurs expériences et de leurs réseaux. Toutefois, le chercheur en charge des nouvelles enquêtes doit réaliser son choix en fonction de sa question de recherche.

Les enquêteurs doivent être sélectionnés en fonction de leur approche disciplinaire et/ou de leur connaissance du terrain. Par exemple, pour réaliser des enquêtes sur l'utilisation de pesticides, on privilégiera des enquêteurs ayant une formation agronomique; alors que pour réaliser des enquêtes auprès d'entreprises agro-alimentaires sur la diversité de leurs actifs, on privilégiera des enquêteurs ayant une formation en gestion. Il faut donc veiller à l'adéquation entre la formation de l'enquêteur et la thématique de recherche.

Une fois sélectionnés, les enquêteurs doivent impérativement être formés par le chercheur qui réalise les enquêtes. Tous les enquêteurs doivent comprendre l'objectif du questionnaire

et donc l'enjeu sous-jacent. De ce fait, il ne doit y avoir aucune ambiguïté et tous doivent comprendre les questions de la même façon afin d'éviter le «biais de l'enquêteur». Autrement dit, il faut être certain que les réponses apportées par les enquêtés ne soient pas conditionnées par l'enquêteur. Pour ce faire, il est essentiel de réaliser au minimum 3-4 jours de formation où l'objectif de la démarche est explicité et les questions clairement définies. La durée de la formation dépendra du nombre d'enquêteurs.

Dans la mesure du possible, les questionnaires tests seront réalisés pendant cette période de formation afin de discuter directement avec les enquêteurs des premiers questionnaires saisis.

5.1.2. Les superviseurs

Le superviseur est à l'interface entre les enquêteurs et le chercheur porteur de la recherche. Son rôle est clef. Il doit superviser les enquêteurs, vérifier le bon déroulement des enquêtes réalisées et contrôler la qualité des questionnaires saisis. De ce fait, il doit impérativement assister à la formation.

Chaque superviseur aura au maximum 5 enquêteurs sous sa responsabilités. Le chercheur porteur du projet est en charge de superviser les superviseurs.

5.2. Test des questionnaires

5.2.1. Objectif de la phase test

La phase test des questionnaires doit permettre d'en confirmer la clarté, de lever les dernières ambiguïtés et de s'assurer que toutes les questions seront renseignées comme attendu. De ce fait, les individus enquêtés lors de cette phase test ne devront pas être enregistrés dans la base de données finale. Cette phase test engendre des coûts supplémentaires à prendre en compte.

5.2.2. Prise de contact avec les fournisseurs de données officielles

Un élément important de la préparation à la collecte est la prise de contact avec des partenaires institutionnels fournisseurs de données officielles. Cette prise de contact permet de rendre légitime le projet et de s'assurer de l'actualisation éventuelle des données.

Dans des contextes où l'accès à certaines bases de données officielles se fait de façon non officielle, les conditions d'accès doivent être formalisées.

L'accès à des données officielles est d'autant plus important qu'il permet la définition et la justification d'un mode d'échantillonnage. Quelle que soit la méthode d'échantillonnage considérée, il est essentiel de pouvoir indiquer dans quelle mesure l'échantillon constitué est représentatif de la population à l'échelle de la zone étudiée, de la région, du pays...



Ce qu'il faut retenir

Préparation à la collecte

- Sélection des enquêteurs et des superviseurs
 - Recrutement en local ?
 - Connaissances du terrain / savoir s'adapter / qualités relationnelles
 - Formation initiale > recommandations
 - Formation
 - Minimum 4 jours
 - Aller-retour avec les enquêtes test
 - Superviseurs : 1 pour 5 enquêteurs
- Test des questionnaires
 - Population enquêtée ne fait pas partie de l'échantillon
 - Planification des enquêtes ?
- Prise de contact avec les fournisseurs de données officielles
 - Légitimer la recherche
 - Positionner l'échantillon

6. Collecte des données

La collecte des informations renvoie à l'organisation, à la planification des enquêtes ainsi qu'à la qualité de la collecte et à l'incitation que peuvent avoir les enquêteurs à adopter un comportement opportuniste.

6.1. Prise de contacts pour faciliter l'accès au terrain

Dans certains cas, il est préférable d'introduire l'enquête grâce à des intermédiaires reconnus par les enquêtés pour favoriser leur participation. Les intermédiaires peuvent être notamment des organismes ou des institutions.

6.2. Planification des rendez-vous

Certaines enquêtes nécessitent une prise de rendez-vous préalable avec les enquêtés, notamment lorsque ceux-ci sont dispersés géographiquement. Cette planification permet de réduire les coûts de déplacement et la durée des enquêtes. Dans ce cadre, le chercheur peut privilégier l'envoi d'un courrier. Le questionnaire peut être joint. L'enquête consiste alors en une discussion autour du questionnaire pré-rempli. Cette discussion doit permettre de compléter les informations non renseignées par les enquêtés.

6.3. Qualité de la collecte et mode de rémunération

Les chercheurs ayant réalisé des enquêtes de terrain ont tous une expérience-anecdotique (enquêtes dupliquées, questionnaires mal saisis, surutilisation des questions-filtres, etc ...)

Ces expériences sont parfois liées au mode de rémunération des enquêteurs. Ceux-ci sont généralement payés au nombre de questionnaires réalisés. Pour s'assurer de la qualité du travail, le contrôle de certains questionnaires de façon aléatoire en cours de période d'enquête est préconisé. Elle justifiera le paiement d'une prime additionnelle. Ce contrôle, coûteux en termes de temps, doit être anticipé. La sélection des superviseurs est un élément clef qui peut permettre au chercheur de minimiser ce coût. A noter que le superviseur doit également être contrôlé par le chercheur porteur du projet.

Une alternative à cette prime à la qualité est la rémunération partielle de l'enquêteur. Une première partie est versée au démarrage des enquêtes. Le solde du paiement intervient en fin de travail si celui-ci correspond au cahier des charges.

D'autre part, le nombre de questionnaires journalier doit être limité à 4 questionnaires/jour/enquêteur.



Ce qu'il faut retenir

Collecte des informations

- Rémunération des enquêteurs
 - Nombre de questionnaires
 - Mise en œuvre d'une prime de qualité
- Vérification de questionnaires en cours d'enquête

7. Saisie et validation des données

7.1. Création de bases de données

Les questionnaires peuvent être saisis sur support papier ou support informatique via un masque de saisi qui permet d'autoalimenter la base de données.

Ce second mode de saisi se traduit par un gain de temps qui présente certaines contraintes :

- La première contrainte est financière (achat d'ordinateurs...)
- La seconde contrainte est matérielle. Il faut avoir accès à un système électrique ainsi qu'à un mode de sauvegarde.
- Enfin, la dernière est d'ordre social. Dans les pays où la pauvreté est omniprésente, utiliser un ordinateur pour la saisie des données peut être mal perçue. Le choix du mode de retranscription n'est pas neutre et il doit être pensé en fonction du contexte dans lequel est réalisée l'enquête.

Les questionnaires renseignés sur papier, mode de saisie le plus fréquent, doivent être retranscrits. A l'UMR MOISA, aucun logiciel n'est imposé. Les principaux logiciels utilisés dans notre unité sont : STATA, SAS, EvIEWS, Excel.

Au sein de l'unité, les chercheurs peuvent bénéficier des services de formation continue de leur institution, ainsi que d'une formation académique. En effet, l'unité participe au Master de Recherche « Agriculture, Alimentation et Développement Durable » (A2D2), co-habité entre l'Université Montpellier I et Montpellier SupAgro. Dans ce cadre, les chercheurs de l'unité peuvent assister à des cours de statistiques dans lesquels le logiciel STATA est présenté aux étudiants. Cette formation permet de rendre les chercheurs autonomes en termes de gestion des données et premières analyses statistiques.

7.2. Validation des données

Valider les données revient à vérifier la cohérence des données recueillies : leur unité et l'articulation entre les variables. Par exemple, pour une variable relative à un niveau de consommation, on s'assurera que les enquêteurs ont correctement renseigné cette information et qu'ils ont tous répondu dans la même unité : euros, kilos.... De la même façon, on s'assurera que la consommation totale déclarée est égale à la somme des différentes consommations considérées.



Ce qu'il faut retenir

Validation des données

- Saisie des questionnaires
 - Papier versus masque de saisie
 - Contrainte financière
 - Contrainte matérielle
 - Contrainte d'ordre sociale
- Choix du logiciel pour
 - Saisie
 - Validation et traitements des données

8. Gestion des dossiers relatifs aux étapes du processus du cycle de vie des données

La démarche doit permettre de justifier de données fiables, traçables, reproductibles, transférables et pérennes. La gestion des différents dossiers relatifs aux étapes du cycle de vie des données s'inscrit dans cette démarche. Pour mettre en œuvre cette gestion, il faut organiser une arborescence qui permette d'identifier rapidement les bases de données, les programmes réalisés ainsi que les résultats obtenus.

La recommandation principale de cette partie est l'adoption d'une codification normalisée pour la dénomination et l'identification des versions des fichiers.

Sur la base des expériences de gestion de ces fichiers par les chercheurs du groupe, une arborescence commune a été validée et est proposée aux chercheurs.



8.1. Bases de données

L'arborescence retenue permet de stocker dans des fichiers distincts, les bases de données initiales des bases de données modifiées. Cette différenciation est indispensable pour ne pas écraser par erreur les fichiers initiaux. Ces fichiers correspondent à la base de données primaire créée lors de la saisie par masque ou à la retranscription des questionnaires si la saisie a été faite sur papier. Dans ce cas, il faut créer un fichier qui contienne les questionnaires originaux scannés, constituant une première sauvegarde. Pour des raisons de lisibilité, il est recommandé de remplir les questionnaires au stylo bleu. Le scan des questionnaires est une sauvegarde supplémentaire. Il limite le risque de perte et évite leur

envoi, parfois coûteux, depuis le terrain d'étude jusqu'à l'unité. Cette démarche est préconisée pour répondre à l'exigence de traçabilité.

8.2. Programme

Une codification normalisée doit être adoptée pour l'enregistrement des différents fichiers de programmation et des variables. Dans le cas où un chercheur travaille en interaction avec d'autres, l'utilisation d'un espace partagé commun est conseillée.

Cela suppose la définition :

- de la codification standardisée pour le groupe ;
- des droits octroyés à chacun (écriture, lecture seule ...) ;
- des accès accordés (un ou plusieurs workpackage).

Le chercheur porteur du projet doit être le garant de la gestion de cet espace partagé.

L'utilisation de fichiers et de variables temporaires permet une gestion optimisée des données, souvent mise en œuvre à des fins de contrôle ou d'étapes intermédiaires. Cette utilisation présuppose une sauvegarde dans un fichier de programmation.

Cette organisation permet la traçabilité des résultats obtenus et présentés dans le cadre scientifique.



Ce qu'il faut retenir

Gestion des dossiers relatifs aux étapes du processus du cycle de vie des données

- Définition d'une arborescence qui permette de répondre à l'objectif de traçabilité des données
 - Bases de données
 - Programmes
 - Sorties
- Réflexion à mener sur variables et fichiers temporaires

9. Archivage

Alors que la sauvegarde correspond à une gestion de données en cours de cycle de vie, l'archivage permet la conservation des données stables, une fois le cycle terminé.

Doivent être archivés :

- Les questionnaires ;
- Les bases initiales ;
- Les bases finales ;
- Les programmes (en précisant notamment la version du logiciel utilisé) ;
- Les articles et publications issus de la recherche ;
- La « carte d'identité ».

Pour la réalisation de la « carte d'identité » de la ou des base(s) de données, il est recommandé de s'appuyer sur la norme internationale ouverte Dublin Core (<http://dublincore.org>). Dans la mesure où la présente charte est applicable uniquement aux bases de données issues d'enquêtes de terrain, les 15 éléments identifiés ne sont pas directement transposables ou pertinents. Le tableau ci-dessous propose une lecture de ces éléments au regard de la spécificité des bases de données considérées.

Propriété	Description	Applicabilité aux données issues d'enquêtes de terrain
Titre	Titre du document : le titre est généralement le nom formel sous lequel la ressource est connue	Nom du projet dans le cadre duquel la ou les bases de données ont été réalisées
Créateur	Créateur du document : nom de la personne, de l'organisation ou du service qui a créé/rédigé le document (auteur et éventuels co-auteurs)	Référent responsable des bases de données (nom + institution + affiliation + email)
Sujet	Le sujet est généralement exprimé sous forme de mots clés, expressions ou codes de classement.	Mots clés.
Description	La description du document peut être composée de : résumé, table des matières, texte libre, représentation graphique.	rédaction d'un "lisez-moi.doc".
Editeur	Il s'agit d'indiquer le nom de la personne, de l'organisation ou du service à l'origine de la publication du document (maison d'édition)	
Contributeur	Il s'agit d'indiquer le nom de la (des) personne(s), de l'organisation ou du service qui a contribué à l'élaboration du document.	Personnes impliquées dans la réalisation et l'exploitation des bases de données (nom + institution + affiliations + email)
Date	Date d'un évènement dans le cycle de vie du document. Généralement il s'agit d'indiquer la date de la création ou la date de la mise à disposition du document.	Date de réalisation du projet : Date de collecte des données Date de mise à disposition des données :
Type	Nature ou genre du contenu ; il décrit les grandes catégories de document. Il est recommandé d'employer un vocabulaire clairement défini tel que le propose le Vocabulaire DCMI Type	Glossaire et identification de chacun des fichiers (leur nom, leur contenu...)
Format	Format du document : physique ou électronique. L'élément Format peut spécifier le matériel et le logiciel utilisés. On peut se référer à la liste des types MIME (Multipurpose Internet Mail Extensions) que propose Dublin Core	Nom et version du logiciel utilisé pour la réalisation des enquêtes et le traitement des données ; format
Identifiant	C'est la référence univoque attribuée (codification au moyen d'un ensemble de caractères et de chiffres). Il est recommandé d'utiliser un système de référencement précis	Numéro d'identification
Source	Référence à une ressource dont la ressource décrite est dérivée. Le document peut en effet découler en totalité ou en partie de la ressource en question.	Harmonisation pour la citation des sources (lors des sorties statistiques ou économétriques ; dans les références bibliographiques...)
Langue	Il s'agit de la langue du document ; il est recommandé d'utiliser une des valeurs définies dans les normes RFC 3066 et/ou ISO 639.	Codes ISO langues : FRE : Français ; ENG : Anglais ; SPA : Espagnol ; GER : Allemand, ITA : Italien
Relation	Lien avec une source apparentée. Il est recommandé d'identifier la ressource apparentée par une chaîne de caractères ou d'un n° conforme à un système d'identification officiel	
Couverture	Périmètre ou domaine d'application du contenu. Il s'agit de la portée du document telle que position géographique, laps de temps, ... Il est recommandé d'utiliser des valeurs dans un vocabulaire normé tel que TGN Thesaurus of Geographic Names (dictionnaire de noms de lieux)	Couverture produit : Couverture spatiale : Continent, Pays, Région, Ville Couverture temporelle : année, période, périodicité
Gestion des droits	Informations sur les droits associés à la ressource. Généralement, l'élément gestion des droits contient une mention de la gestion des droits afférents à la ressource ou une référence au service fournissant cette information. L'information sur les droits englobe souvent les droits de propriété intellectuelle (IPR), le copyright et divers droits de propriété.	Règles contractuelles relatives aux droits d'utilisation des données mises à disposition

Au-delà des éléments identifiés *via* la norme Dublin Core, la spécificité des données issues d'enquêtes de terrain nécessite de prendre en compte et de mentionner les éléments suivants :

- Commanditaire(s) du projet
- Institutions impliquées
- Mention sur la mise à disposition des données et ses conditions

Cette bibliothèque ainsi constituée est le reflet et la mémoire des enquêtes de terrain réalisées dans l'UMR.



Ce qu'il faut retenir

Archivage

- Fiche d'identité de chaque base
- Réflexion sur le lieu de stockage

10. Accès et diffusion des données

10.1. Anonymisation

Garantir aux enquêtés de ne pas révéler leur identité est une condition parfois nécessaire à la réalisation d'enquête. Il est indispensable de le stipuler explicitement en début de questionnaire que les données seront traitées individuellement mais que les résultats obtenus seront diffusés de manière agrégée.

Le chercheur porteur de la recherche doit être le garant de cette anonymisation. Il doit créer d'un fichier dans lequel il dispose de l'identité de l'enquêté et du numéro d'identification. Les fichiers qu'il échangera avec les partenaires ou autres chercheurs éventuels doivent être les fichiers anonymisés. Le fichier permettant l'identification ne doit pas être diffusé.

10.2. Confidentialité des données au sein du groupe de travail

La confidentialité des données est à définir dès la rédaction du contrat. Les règles de confidentialité doivent impliquer tous les acteurs: bailleurs, partenaires, chercheurs, doctorants, etc... A cette occasion, l'ensemble des partenaires pourra être amené à s'engager sur un cahier des charges précisant, comme le contrat les conditions:

- d'accès aux données
- d'utilisation des données
- de conservation des données
- de stockage
- de citation des sources (données, projet, etc ...)
- de modalités de diffusion des résultats

10.3. Déclaration CNIL

Lors du recueil de données individuelles non anonymisées, il est obligatoire de faire une déclaration auprès de la CNIL (Commission Nationale Informatique et Libertés) <http://www.cnil.fr/vos-obligations/declarer-a-la-cnil/> .

Tout recueil de données individuelles doit faire l'objet d'une déclaration dès lors que les individus enquêtés peuvent être identifiés, directement ou indirectement, à partir des réponses apportées au questionnaire.

La déclaration CNIL a pour objectif de protéger les individus. Elle est définie dans le cadre de la loi du 6 Janvier 1978. Plus précisément, 2 articles nous concernent:

- article 4: «... sont réputées nominatives... les informations qui permettent sous quelque forme que ce soit, directement ou non, l'identification des personnes physiques auxquelles elles s'appliquent»
- article 5: «... traitement automatisé d'informations nominatives... tout ensemble d'opérations réalisées par les moyens automatiques, relatif à la collecte, l'enregistrement, l'élaboration, la modification, la conservation et la destruction d'informations nominatives ainsi que tout ensemble d'opérations de même nature se rapportant à l'exploitation de fichiers ou bases de données et notamment les interconnexions ou rapprochement, consultations ou communications d'informations nominatives ».

Dans le cadre de données nominatives qui répondent aux articles ci-dessus, la procédure à suivre est la suivante :

- remplir une déclaration en fournissant l'ensemble des pièces exigées
- adresser la déclaration en 3 exemplaires en recommandé avec accusé réception
- attendre le récépissé avec numéro d'enregistrement

Les formalités et obligations :

- préciser dans le document la finalité de la recherche, la nature des données collectées, la durée d'utilisation, les personnes qui auront accès aux données
- ne pas détourner la finalité de la recherche
- ne pas conserver les données au-delà de la période prévue
- veiller à la sécurité des données
- veiller à ce qu'un tiers ne puisse pas détourner les données
- ne pas communiquer les informations à des personnes non autorisées et identifiées dans la déclaration
- ne pas commercialiser les données recueillies hors d'un cadre légal
- se conformer au droit français

Il est important de souligner que le responsable du traitement est juridiquement responsable du contenu de la déclaration CNIL.

10.4. Mise à disposition

La mise à disposition des données, lorsqu'elle est possible, est à l'initiative du chercheur. L'accord devra être formalisé et stipulera les conditions d'accès : respect de l'anonymat, citation du chercheur, de son unité de rattachement, de son institution, des sources, ... La « carte d'identité » précise l'ensemble de ces éléments.



Ce qu'il faut retenir

Mise à disposition & droits de propriété

- Anonymisation des données
 - Identification d'un garant
 - Création d'un fichier identifiant/numéro anonymisation
- Mise à disposition, droit de propriété et partenariats
 - Démarche individuelle
 - Définir les règles :
 - ex-ante
 - par écrit